Parsing Approaches for Swiss German

Aepli, Noëmi ; Clematide, Simon

Parsing Approaches for Swiss German

Noëmi Aepli
University of Zurich
noemi.aepli@uzh.ch

Simon Clematide
University of Zurich
simon.clematide@cl.uzh.ch

Abstract

This paper presents different approaches towards universal dependency parsing for Swiss German. Dealing with dialects is a challenging task in Natural Language Processing because of the huge linguistic variability, which is partly due to the lack of standard spelling rules. Building a statistical parser requires expensive resources which are only available for a few dozen high-resourced languages. In order to overcome the low-resource problem for dialects, approaches to cross-lingual learning are exploited. We apply different cross-lingual parsing strategies to Swiss German, making use of Standard German resources. The methods applied are *annotation projection* and *model transfer*. The results show around 60% Labelled Attachment Score for all approaches and provide a first substantial step towards Swiss German dependency parsing. The resources are available for further research on NLP applications for Swiss German dialects.

1 Introduction

Swiss German is a dialect continuum of the Alemannic dialect group, comprising numerous varieties used in the German-speaking part of Switzerland.[1] Unlike other dialect situations, the Swiss German dialects are deeply rooted in the Swiss culture and enjoy a high reputation, i.e. dialect speakers are not considered less educated as it is the case in other countries. On the basis of their high acceptance in the Swiss culture and with the introduction of digital communication, Swiss German has undergone a spread over all kinds of communication forms and social media. Despite being oral languages, the dialects are used increasingly in written contexts, and writers spell as they please.

For Natural Language Processing (NLP), low-resourced languages are challenging, particularly in cases like Swiss German where no orthographic rules are followed. Compiling NLP resources from scratch such as syntactically annotated text corpora (treebanks) is a laborious and expensive process. Thus, in such cases, cross-lingual approaches offer a perspective to get started with automatic processing of the respective language. Such approaches are especially promising if a closely related resource-rich language is available, which is the case for Swiss German.

The *Universal Dependencies* (*UD*) project aims at developing and setting a standard for cross-linguistically consistently annotated treebanks in order to facilitate multilingual parsing research. We support this idea by adopting the current UD standard as much as possible.

The information about which word of the sentence is dependent on which other one is important in order to correctly understand the meaning of a sentence. Thus, it is needed for numerous NLP applications like information extraction or grammar checking. The task of identifying these dependencies is done by a dependency parser (see Figure 1 for a Swiss German example in UD).

In this paper, we apply two different cross-lingual dependency parsing strategies, namely annotation projection as a lexicalised approach, and model transfer as a delexicalised approach. We manually create a gold standard in order to evaluate and compare the different strategies. Furthermore, we build and evaluate

In: Mark Cieliebak, Don Tuggener and Fernando Benites (eds.): Proceedings of the 3rd Swiss Text Analytics Conference (SwissText 2018), Winterthur, Switzerland, June 2018

[1] Swiss Standard German, one of the four official languages of Switzerland, is not to be confused with Swiss German dialects.

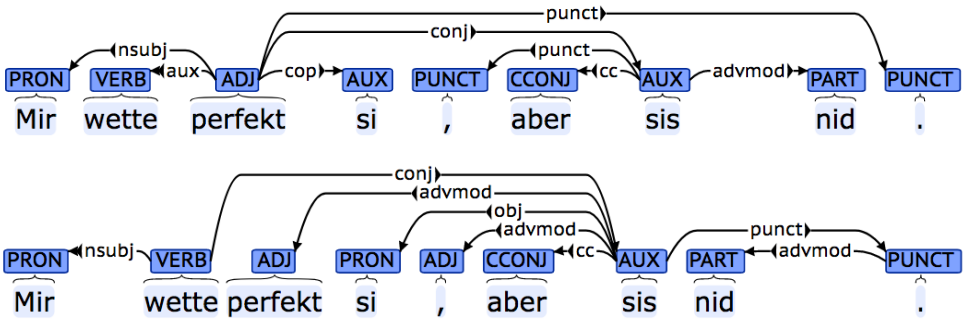

Figure 1: Universal dependency parse trees for the sentence: *We want to be perfect, but we are not.* Top: gold standard, bottom: system.

a *silver standard treebank* which, compared to manually annotating from scratch, accelerates the creation of a larger training set for a monolingual Swiss German parser.

The next section presents related work on NLP for Swiss German and introduces the two main approaches to cross-lingual parsing. In Section 3 and 4 we present our data and methods. Section 5 shows and discusses our results.

2 Related Work

Even though there have been several projects involving Swiss German (Hollenstein and Aepli, 2014; Zampieri et al., 2017; Hollenstein and Aepli, 2015; Samardzic et al., 2016; Samardžić et al., 2015; Scherrer, 2007; Baumgartner, 2016; Dürscheid and Stark, 2011; Stark et al., 2014; Scherrer and Owen, 2010; Scherrer, 2013, 2012), resources for NLP applications are still rare. As so often for dialects, even data for Swiss German is sparse. Therefore, the approach is to use tools and data of related resource-rich languages and apply transfer methods.

2.1 Universal Dependencies

Research in dependency parsing has increased significantly since a collection of dependency treebanks has become available, in particular through the *CoNLL shared tasks on dependency parsing* (Buchholz and Marsi, 2006; Nivre et al., 2007a; Zeman et al., 2017) which have provided many data sets. In order to facilitate cross-lingual research on syntactic structure and to standardise best-practices, *Universal POS* (*UPOS*) tags (Petrov et al., 2012) as well as *Universal Dependencies* (Nivre et al., 2016) have been introduced. The annotation scheme is originally based on *Stanford dependencies* (de Marneffe et al., 2006; de Marneffe and Manning, 2008; de Marneffe et al., 2014). McDonald et al. (2013) present the first collection of six treebanks with homogenous syntactic dependency annotation, which has continually been expanded since.

2.2 Cross-lingual Dependency Parsing

There are two main approaches to cross-lingual syntactic dependency parsing. Firstly, the delexicalized *model transfer* of which the goal is to abstract away from language-specific parameters, i.e. train delexicalised parsers. The idea is based on universal features and model parameters that can be transferred between related languages. Hence, this method assumes a common feature representation across languages. The advantage of the model transfer approach is that no parallel data is needed. Zeman and Resnik (2008) train a basic delexicalised parser relying on part-of-speech (POS) tags only. McDonald et al. (2013); Petrov et al. (2012) and Naseem et al. (2010) rely on universal features while Täckström et al. (2013) adapt model parameters to the target language in order to cross-linguistically transfer syntactic dependency parses.

The main idea of the second approach, the lexicalised *annotation projection* method, is the mapping of labels across languages using parallel sentences and automatic alignment. It includes projection heuristics and usually post projection rules. The main drawback of this approach is that it relies on sentence-aligned parallel corpora. In order to deal with this restriction, treebank translation has emerged where the training data is automatically translated with a machine translation system. The central point of this method is the alignment along which the annotations are mapped from one language to the other. Automatic word alignment has already been used by Yarowsky et al.

(2001); Aepli et al. (2014) and Snyder et al. (2008) for improving resources and tools for POS tagging of supervised and unsupervised learning respectively. Hwa et al. (2005), Tiedemann (2014) and Tiedemann (2015) use annotation projection approaches for parsing, and Tiedemann et al. (2014) as well as Rosa et al. (2017) use machine translation in addition instead of relying on parallel corpora. For Swiss German, treebank translation is not viable because of sparse data and the lack of a Machine Translation system for Swiss German. Hence, in this paper we apply annotation projection as a lexicalised approach and model transfer as a delexicalised approach.

3 Materials

3.1 Standard German Data

We use the *German Universal Dependency treebank*[2] consisting of 13,814 sentences. It is annotated according to the *UD* guidelines[3] and contains *Universal POS* (*UPOS*) tags (Petrov et al., 2012). The treebank comes in `CoNLL-U` format but as some tools cannot handle it, we convert it to `CoNLL-X`. This includes one major tokenization change concerning the *Stuttgart-Tübingen-TagSet* (*STTS*) (Schiller et al., 1999) POS tag `APPRART`. In `CoNLL-U` the prepositions with fused articles are split into two syntactical words. We undo this split, merge the information in one token and correspondingly adapt the dependency relations.

3.2 Swiss German Data

Annotation projection requires a parallel corpus. The *AGORA citizen linguistics* project[4] crowdsourced Standard German translations of 6,197 Swiss German sentences via the web site *dindialaekt.ch*. The sentences are taken from the *NOAH* corpus (Hollenstein and Aepli, 2014), additionally, sentences from novels in Bernese and St Gallen dialect were added to better represent syntactic word order differences. By the end of November 2017, the citizen linguists produced 41,670 translations. We aggregated and cleaned the data into a parallel GSW/DE corpus of 26,015 sentences. In particular, we filtered translations that differed too much in length or Levenshtein edit distance[5] from the Swiss German source sentence.

[2] https://github.com/UniversalDependencies/UD_German
[3] http://universaldependencies.org/guidelines.html
[4] https://www.linguistik.uzh.ch/de/forschung/agora.html

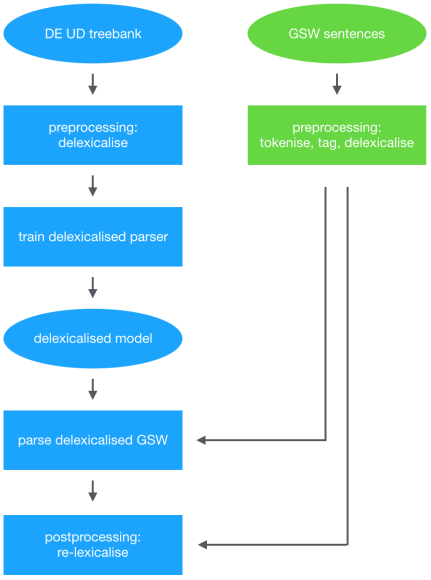

Figure 2: Workflow of the model transfer.

4 Methods

We apply two classical parsing approaches presented in Section 2: model transfer with a delexicalised parser and annotation projection with crowdsourced parallel data. Within both approaches we test two parsing frameworks; the *MaltParser* (Nivre et al., 2007b) and the more recent *UDPipe* (Straka and Straková, 2017). Both parsers are provided with tokenised input.

4.1 Model Transfer Approach

The delexicalised model transfer approach is straightforward, working on the basis of POS tags only. For the training, the words in the Standard German corpus are replaced by their POS tags. Accordingly, at parsing time the Swiss German words are replaced by their POS tag before parsing and re-inserted afterwards.

4.1.1 POS tagging

Part-of-speech tagging is an important step prior to parsing because the syntactic structure builds upon the

[5] The Levenshtein distance (Levenshtein, 1966) measures the difference between two sequences of characters. Hence, the minimal edit distance between two words is the minimum number of characters to be changed (i.e. inserted, deleted or substituted), in order to make them equal.

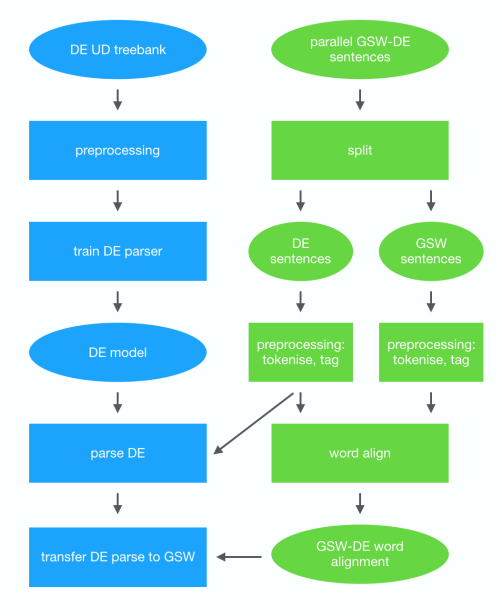

Figure 3: Workflow of the annotation projection.

POS information. Obviously, when training delexicalised parsers, this step is crucial as the tags are the only information available to the parser.

For POS tagging Swiss German sentences, we used the *Wapiti* (Lavergne et al., 2010) model trained on Release 2.2 of the *NOAH* corpus, where average accuracy in 10-fold crossvalidation is 92.25%.

The `CoNLL`-format includes *UPOS* tags in addition to the fine-grained language-specific POS tags (*STTS* in the case of German and Swiss German). We used the mapping provided by the *UD* project in order to infer the *UPOS* tags from the given *STTS* tags.

4.2 Annotation Projection Approach

Annotation projection is not only more complex in processing compared to model transfer but also needs more resources. Most importantly, annotation projection requires a word-aligned parallel corpus. Starting from the crowdsourced sentences which are sentence-aligned, it is the task of a word aligner to compute the most probable word alignments, i.e. the information about which word of the (Swiss German) source sentence corresponds to which word of the target sentence, i.e. the translation. There are many tools for this as it is a basic step also in machine translation systems. We tested three of them: *GIZA++* (Och and Ney, 2003), *FastAlign* (Dyer et al., 2013) and *Monolingual Greedy Aligner* (MGA) (Rosa et al., 2012).

The idea of the annotation projection process is to use the tool (here: parser) of a resource-rich language on that language (here: German) and then project the generated information (here: universal dependency structures) along the word alignment to the target language (here: Swiss German). In practice, this means we train the parser on the Standard German treebank (see Section 3.1) and parse the Standard German translations of the Swiss German original sentences. Then we project the resulting parse structure along the word alignments from the German word to the corresponding Swiss German word.

4.2.1 Transfer of the Annotation

The transfer is the core component of annotation projection. The parse of the Standard German translation is projected along the word alignment to its Swiss German correspondent. The input consists of the Standard German parse and the alignment between the Standard German sentence and its Swiss German version (`GSW:DE`). Algorithm 1 describes the projection process.

Data: DE parse & alignment `GSW:DE`
Result: DE parse transferred to GSW
for *word alignment in sentence* **do**
  if *`1:1` alignment* **then**
    transfer parse of DE
  else if *`1:0` alignment (i.e. no DE word aligned)* **then**
    attach GSW word to `root` as POS tag `ADV` and dependency label `advmod`
  else *`1:n` alignment (i.e. several DE words aligned)*
    transfer parse of aligned DE word with smallest edit (Levenshtein) distance
  end
end

Algorithm 1: Transfer of parses.

The case of `1:1` alignment where exactly one German word is aligned to the Swiss German word is easy; the only thing to do is projecting the dependency of the German word to the Swiss German word. If, however, there are several German words aligned to one Swiss German word (`1:n`), the algorithm has to decide which parse to transfer.[6] In order to take this decision, the algorithm computes the Levenshtein

[6]Note that the case of a `1:n` alignment between a GSW word and a DE multiword expression is not covered in this approach.

distances (Levenshtein, 1966) between the Swiss German word and every aligned German word and takes the one with the smallest edit distance. The most challenging case is when no German word is aligned to the Swiss German token. A simple baseline approach attaches the corresponding Swiss German word as adverbial modifier to the root of the sentence.

The decision to treat every unaligned Swiss German word as an adverb is taken on the basis of the frequency distribution of POS tags; ADV is the second most frequent POS tag (after NN) in the Swiss German data. However, taking into consideration the word itself, some more sophisticated rules can be elaborated. Considering the differences between Standard German and Swiss German as described by Hollenstein and Aepli (2014), we can expect some words like infinitive particles (PTKINF) (e.g. *go*) or the past participle *gsi* (*been*) to remain unaligned. The former because these words do not exist in Standard German, the latter because Standard German simple past tense is expressed by perfect tense in Swiss German, typically resulting in a "spare" past participle in the alignment. Furthermore, there are unaligned articles because Swiss German requires articles in front of proper names. Also punctuation including the apostrophe is a source of errors which can easily be corrected. The application of these more elaborate rules have an impact of around 2 points on the evaluation scores.

Algorithm 1 transfers the German parses as they are, as a consequence the numbering of the token IDs is mixed up. Correcting the token IDs to be in ascending order (from 1 to the length of the sentence) requires the corresponding adjustment of the head references. Furthermore, one needs to make sure that there is exactly one root in a sentence.

Data: transferred DE parse to GSW words
Result: valid GSW parse
for *sentence in parse* **do**
  if *DE root was not projected to GSW parse* **then**
    take 1st VERB as root, else 1st NOUN
  else if *head of a projected word was not projected to GSW parse* **then**
    attach it to the root
  end
end

Algorithm 2: Correction of transferred parses.

Algorithm 2 goes through every sentence of the input file and first makes sure that there is one root for the sentence. If the root of the Standard German parse has not been transferred to the Swiss German sentence (missing word alignment), the first verb (*UPOS* VERB) is taken as root and if there is no VERB in the sentence, the first NOUN is considered the root.

4.3 Optimisation

We tested two approaches for optimisation; preprocessing of the training set and postprocessing rules to be applied after parsing.

4.3.1 Preprocessing

One frequent mistake mostly observed in the delexicalised approach is the wrong assignment of passive dependency labels instead of their active counterpart. The passive construction in Standard German is built with the auxiliary *werden*, which can, however, also be used in non-passive constructions. The combination of VA* and a perfect participle (VVPP) is very frequent in Swiss German, however, it is usually not a passive construction but rather a perfect tense. Therefore, a simple but effective solution is the introduction of a new "set" of POS tags in the German *UD* training set: VWFIN, VWINF, VWPP for finite verbs, infinitives and participles respectively of the verb *werden*. This means, all occurrences of the lemma *werden* as an auxiliary (i.e. *UPOS*: AUX and *STTS*: V**A**{INF|PP}) are replaced by V**W**{INF|PP}. In this way, the system learns to discriminate between the usage of *werden* as auxiliary versus the usage as full verb and, most of all, it learned to differentiate between the auxiliary *werden* and the other auxiliaries *haben* (*to have*) and *sein* (*to be*). Hence, the number of wrongly assigned passive dependency labels decreased, which leads to an improvement of around 2.5 to 3.5 points as presented in Section 5.

4.3.2 Postprocessing

Some of the errors can easily be corrected with simple rules in a postprocessing step. One example is a frequent error caused by a remnant of the 1st UD version which is handled differently in *UD version 2*. The two labels oblique nominal (obl) and nominal modifier (nmod) are confused because the latter was used to modify nominals and predicates in *UD v1*. However, in *UD v2*, obl is used for a nominal functioning

as an oblique argument, while nmod is used for nominal dependents of another noun (phrase) only. This means, if the head is a verb, adjective or adverb, the dependency label has to be obl. If, instead, the head is a noun, pronoun, name or number, the dependency label is nmod.

5 Results & Discussion

This section presents the different settings and combinations of aforementioned resources, approaches and tools. For the evaluation, we manually created a gold standard consisting of 100 Swiss German sentences taken from the resources presented in Section 3.2. We evaluated the approaches according to Labelled Attachment Score (LAS) and Unlabelled Attachment Score (UAS)[7], not excluding punctuation. The results we present here are macro accuracy scores, that is, the scores are computed separately for each sentence and then averaged[8]. Note that there is a mismatch in the actual annotation of punctuation between the the Standard German *UD* treebank v2 and the official guidelines we were applying. This difference in the punctuation dependencies has an effect on the scores, i.e. it lowers the scores presented here. Furthermore, note that the test set containing 100 gold standard sentences is small and therefore these results have to be taken with a grain of salt.

5.1 German Parser Accuracy

In order to put the results into context, we checked the performance of the parsers on the German *UD v2* treebank using their split of training and test set. In this setting, we left all the available information for the parser to use, including morphology and lemmas. The APPRART splitting is undone for the CoNLL-X *MaltParser* input, not so for the *UDPipe* which takes CoNLL-U as input format (and performs worse with the *MaltParser*-CoNLL-X input). *MaltParser* reaches a LAS of 79.71%, *UDPipe* 70.31% respectively.

5.2 Direct Cross-lingual Parsing

As a comparison to the main approaches, we applied Standard German parsers directly to Swiss German. This means, we used the training set of the German *UD* treebank to train the *MaltParser* (using *MaltOptimizer* to get the best hyperparameter settings) and *UDPipe*. Before training, we removed the morphology and lemma information because this information is not available in the Swiss German test set and therefore the parsers cannot rely on it. Furthermore, for the *MaltParser* we converted the training set from CoNLL-U to CoNLL-X format because *MaltOptimizer* cannot handle the former. Testing the *MaltParser* model on the gold standard with automatically assigned POS tags by *Wapiti* results in an LAS of 55.28%. *UDPipe* only reaches 21.19% LAS, one reason for this low accuracy could be that *UDPipe* relies on word embedding information (Straka and Straková, 2017), which results in a low recall when applying a model trained on German to Swiss German.

5.3 Delexicalised Model Transfer

Instead of giving the parser the Standard German words as input like in the direct cross-lingual approach, in the delexicalised approach we provide the parser with POS information only. This means, the words are replaced by *STTS* POS tags while all the other columns stay the same. Given the small evaluation set and a negligible difference in the results, the two parsers' performance can be considered the same: $\sim$57% LAS for both when trained on the preprocessed training set, i.e. differentiating the auxiliary *werden* vs. the auxiliaries *haben* (*to have*) and *sein* (*to be*) (see Section 4.3.1).

5.4 Annotation Projection

The results for the annotation projection approach vary substantially depending on the combination of aligner and parser. Starting from 46.45% LAS (*MaltParser* + *Fastalign*), the combination of *UDPipe* and *Monolingual Greedy Aligner* scores best in this approach with 53.39% LAS. This score is reached with the baseline transfer rules where unaligned words are simply attached to the root as adverbs. Applying more elaborate transfer rules (Section 4.2.1) results in an improvement of 2.09 points to 55.65% LAS. The preprocessing step does not improve the results in this approach. These results show that the *Monolingual Greedy Aligner* performs best in the task of DE/GSW alignment. *MGA* takes character-based word similar-

[7] UAS is the percentage of tokens with the correct syntactic head, LAS the percentage of tokens assigned the correct syntactic head as well as the correct dependency label.

[8] Macro accuracy scores as opposed to the word-based micro scores, where the true positives are summed up over the whole treebank and divided by the total number of words.

ity into account which intuitively makes sense as the information about similar letters is valuable information when dealing with closely related languages such as Standard German and Swiss German.

5.5 Postprocessing

The postprocessing rules do not show a huge impact on the parsing results; the `nmod`/`obl` confusions for example are still present. The reason for this is that the parser assigned wrong heads to many of the words and therefore the rule to correct the `nmod`/`obl` confusions does not work. The LAS scores improve by 1.62 points for the cross-lingual *MaltParser* and 2.07 points for delexicalised model transfer and annotation projection *UDPipe* approachs respectively, reaching nearly **60%** LAS accuracy.

5.6 Discussion

Table 1 shows the best results including the corresponding setting for every approach. The best LAS results of all the applied approaches are very close, hence there is no clear answer to the question of which approach works best. Annotation projection is the most laborious among the three and as such not the first option to choose. Furthermore, the transfer of the annotation is strongly dependent on the performance of the aligner, which in turn benefits from big parallel corpora to be trained on. However, such big parallel corpora do not exist yet for Swiss German dialects.

Contrary to our expectations, training specific models for different dialects does not have a huge impact on the results. The word ordering for the St Gallen dialect is closer to the Standard German word ordering while Bernese dialect speakers often change the order of the verbs. Due to these differences, we expected the model transfer approach to perform worse on the Bern dialect than annotation projection, where the word order changes should be handled by the aligner. Looking specifically at Bernese sentences with "switched" word order (e.g. *ha aafo gränne* ('I started to cry'), *gfunge hei gha* ('have found'), *het übercho* ('have gotten')), there is no significant difference between the two approaches in our test set.

5.6.1 Swiss German Variability

The results presented here are not perfect and certainly require further improvement in order for a system to be used in real-life applications. Compared with the Standard German parser accuracy, which reaches almost 80% LAS on the German *UD v2* with standard settings of the parsers, there is room for improvement. However, these numbers have to be set in relation to the data we worked with. Even though we could make use of Swiss German novels and crowdsourced data, it is still a small data set. Furthermore, the enormous spelling variability in Swiss German dialects poses a serious challenge for all tools. Statistical tools work best if the observed events are frequent. However, they do not work well with sparse data consisting of a large amount of hapax legomena, i.e. word form which appears only once. Figure 4 shows the frequencies (on y-axis) of type frequencies (x) in a Swiss German text collection[9] consisting of 6,155 sentences with 105,692 tokens and 20,882 unique token types. 14,099 types appear only once (i.e. hapax legomena), 2,804 appear twice (i.e. hapax dislegomena) 19,874 less than 10 times and 20,767 less than 100 times.

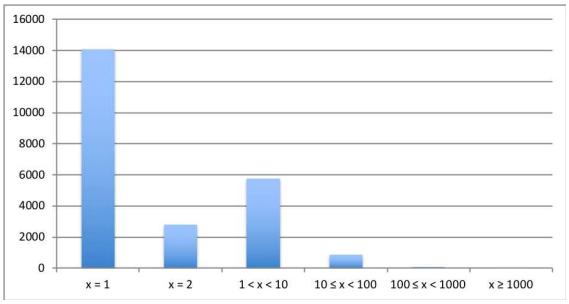

Figure 4: Frequencies of type frequencies (x) in a Swiss German text.

5.7 Silver Treebank Parsing Model

Following the direct cross-lingual parsing approach, we automatically parse 6,155 Swiss German sentences[9] in order to create a *silver treebank*. A silver standard treebank, as opposed to a gold standard treebank which is assumed to be correctly annotated, is automatically annotated and may therefore contain errors. Then, we use this silver treebank to train a monolingual Swiss German parser and hence, create a first monolingual Swiss German dependency parsing model. The advantage of using a silver treebank is the fact that it becomes a monolingual task. However, this comes with the price of a faulty training set, which is not the best resource to build a parser.

[9] *NOAH* corpus plus 396 sentences from novels by Pedro Lenz and Renato Kaiser, excluding gold standard sentences.

Table 1: Comparison of the best score of every approach.

Approach	Setting	LAS	UAS
Annotation Projection	UDPipe + MGA + Postprocessing	57.73	66.57
Model Transfer	UDPipe + Pre- and postprocessing	**60.64**	**72.48**
Direct Cross-lingual	MaltParser (+ Wapiti) + Pre- and postprocessing	59.78	70.80

Interestingly, the performance of the *MaltParser* trained on the silver treebank reaches the same performance as the direct cross-lingual parsing approach itself, which was used to generate the silver treebank: LAS 57.10%. Given that 6,000 sentences do not constitute a large training set for a statistical parser, a parser could probably profit from additional related Standard German material. However, combining the two training sets, i.e. the German *Universal Dependency* treebank and the silver treebank gives slightly worse results (LAS 55.46%).

5.8 Future Work

There are several opportunities for further improvement. Concerning the annotation projection approach, the crucial alignment information needs to be improved for example by ensembling over results from different word aligners. In cases where alignment does not work, adding further transfer and postprocessing rules would be important. In addition, a spelling normalisation strategy can help to deal with the data sparseness imposed by the phonetic and orthographic variability in Swiss German dialects. Moreover, the outputs of the three parsing approaches could be ensembled, e.g. via majority vote like for alignment as aforementioned, to get rid of the weaknesses of each approach. Furthermore, the *silver treebank* created could be manually corrected in order to generate a treebank which can be used as training set for a monolingual dependency parser for Swiss German. Finally, once the data sparseness for Swiss German varieties is mitigated, modern neural methods are promising as shown for example in the work by Ammar et al. (2016). Ammar et al. train one multilingual model that can be used to parse sentences in several languages. In order to do so, they use many resources including a bilingual dictionary for adding cross-lingual lexical information, and a monolingual corpus for training word embeddings. Such approaches need a big amount of data of the language to be parsed, which are still not available for Swiss German.

6 Conclusion

In this work, we experimented with a variety of cross-lingual approaches for parsing texts written in Swiss German. For statistically driven systems, languages with non-standardised orthography are a demanding task. Swiss German dialects feature challenging Natural Language Processing (NLP) problems with their lack of orthographic spelling rules and a huge pronunciation variety. This is a situation which leads to a high degree of data sparseness and with it, a lack of resources and tools for NLP.

We tested a lexicalised annotation projection method as well as a delexicalised model transfer method. The annotation projection method requires parallel sentences in both the resource-rich and the low-resourced language while the delexicalised model transfer approach only requires a monolingual treebank of a closely related resource-rich language.

The evaluation on a manually annotated gold standard consisting of 100 sentences shows a 60% Labelled Attachment Score (LAS) with negligible differences between the different parsing approaches. However, the annotation projection approach is more complex than model transfer due to the transfer rules and the crucial word alignment process.

This work provides a first substantial step towards closing a big gap in Natural Language Processing tools for Swiss German and provides data[10] to work on further improvements.

Acknowledgments

We thank the *AGORA* project "Citizen Linguistics" for making their translation data available and in particular all our volunteer translators. We also thank Renato Kaiser and Pedro Lenz for their permission to use their novels in our experiments.

[10] https://github.com/noe-eva/SwissGermanUD

References

Noëmi Aepli, Tanja Samardžić, and Ruprecht von Waldenfels. 2014. Part-of-Speech Tag Disambiguation by Cross-Linguistic Majority Vote. In *First Workshop on Applying NLP Tools to Similar Languages, Varieties and Dialects (VarDial)*. Dublin.

Waleed Ammar, George Mulcaire, Miguel Ballesteros, Chris Dyer, and Noah A. Smith. 2016. Many languages, one parser. *TACL* 4:431–444.

Reto Baumgartner. 2016. Morphological analysis and lemmatization for Swiss German using weighted transducers. In Stefanie Dipper, Friedrich Neubarth, and Heike Zinsmeister, editors, *Proceedings of the 13th Conference on Natural Language Processing (KONVENS 2016)*. Bochum, Germany.

Sabine Buchholz and Erwin Marsi. 2006. Conll-x shared task on multilingual dependency parsing. In *In Proceedings of CoNLL*. pages 149–164.

Marie-Catherine de Marneffe, Timothy Dozat, Natalia Silveira, Katri Haverinen, Filip Ginter, Joakim Nivre, and Christopher D. Manning. 2014. Universal stanford dependencies: A cross-linguistic typology. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014)*.

Marie-Catherine de Marneffe, Bill MacCartney, and Christopher D. Manning. 2006. Generating typed dependency parses from phrase structure parses. In *5th International Conference on Language Resources and Evaluation (LREC 2006)*.

Marie-Catherine de Marneffe and Christopher D. Manning. 2008. The Stanford typed dependencies representation. In *COLING Workshop on Cross-framework and Cross-domain Parser Evaluation*.

Christa Dürscheid and Elisabeth Stark. 2011. SMS4science: An international corpus-based texting project and the specific challenges for multilingual Switzerland. *Digital Discourse: Language in the New Media* pages 299–320.

Chris Dyer, Victor Chahuneau, and Noah A. Smith. 2013. A simple, fast, and effective reparameterization of IBM model 2. In *HLT-NAACL*. pages 644–648.

Nora Hollenstein and Noëmi Aepli. 2014. Compilation of a swiss german dialect corpus and its application to pos tagging. In *Proceedings of the First Workshop on Applying NLP Tools to Similar Languages, Varieties and Dialects*. Dublin, Ireland, pages 85–94. http://www.aclweb.org/anthology/W14-5310.

Nora Hollenstein and Noëmi Aepli. 2015. A resource for natural language processing of swiss german dialects. In *Proceedings of the International Conference of the German Society for Computational Linguistics and Language Technology*. German Society for Computational Linguistics and Language Technology, Duisburg-Essen, Germany, pages 108–109.

Rebecca Hwa, Philip Resnik, Amy Weinberg, Clara Cabezas, and Okan Kolak. 2005. Bootstrapping parsers via syntactic projection across parallel texts. *Natural language engineering* 11(03):311–325.

Thomas Lavergne, Olivier Cappé, and François Yvon. 2010. Practical Very Large Scale CRFs. In *Proceedings the 48th Annual Meeting of the Association for Computational Linguistics (ACL)*. Uppsala, Sweden, pages 504–513.

VI Levenshtein. 1966. Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *Soviet Physics Doklady* 10:707.

Ryan McDonald, Joakim Nivre, Yvonne Quirmbach-Brundage, Yoav Goldberg, Dipanjan Das, Kuzman Ganchev, Keith Hall, Slav Petrov, Hao Zhang, Oscar Täckström, Claudia Bedini, Núria Bertomeu Castelló, and Jungmee Lee. 2013. Universal dependency annotation for multilingual parsing. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. pages 92–97.

Tahira Naseem, Harr Chen, Regina Barzilay, and Mark Johnson. 2010. Using universal linguistic knowledge to guide grammar induction. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*. pages 1234–1244.

Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajic, Christopher D. Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty, and Daniel Zeman. 2016. Universal dependencies v1: A multilingual treebank collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. Paris, France.

Joakim Nivre, Johan Hall, Sandra Kübler, Ryan McDonald, Jens Nilsson, Sebastian Riedel, and Deniz Yuret. 2007a. The CoNLL 2007 shared task on dependency parsing. In *Proceedings of the CoNLL Shared Task Session of EMNLP-CoNLL 2007*. Prague, Czech Republic, pages 915–932. http://www.aclweb.org/anthology/D/D07/D07-1096.

Joakim Nivre, Johan Hall, Jens Nilsson, Atanas Chanev, GülŞen Eryigit, Sandra Kübler, Svetoslav Marinov, and Erwin Marsi. 2007b. Maltparser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering* 13(2):95–135. https://doi.org/10.1017/S1351324906004505.

Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics* 29(1):19–51.

Slav Petrov, Dipanjan Das, and Ryan McDonald. 2012. A universal part-of-speech tagset. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*. pages 2089–2096.

Rudolf Rosa, Ondřej Dušek, David Mareček, and Martin Popel. 2012. Using parallel features in parsing of machine-translated sentences for correction of grammatical errors. In *Proceedings of the Sixth Workshop on Syntax, Semantics and Structure in Statistical Translation*. Jeju, Republic of Korea, pages 39–48. http://www.aclweb.org/anthology/W12-4205.

Rudolf Rosa, Daniel Zeman, David Mareček, and Zdenek Žabokrtský. 2017. Slavic forest, Norwegian wood. In *Proceedings of the Fourth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial)*. Valencia, Spain, pages 210–219. http://www.aclweb.org/anthology/W17-1226.

Tanja Samardžić, Yves Scherrer, and Elvira Glaser. 2015. Normalising orthographic and dialectal variants for the automatic processing of Swiss German. In *Language and Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics*. Poznan, Poland, pages 294–298.

Tanja Samardzic, Yves Scherrer, and Elvira Glaser. 2016. ArchiMob – a corpus of spoken Swiss German. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. Paris, France.

Yves Scherrer. 2007. Adaptive string distance measures for bilingual dialect lexicon induction. In *Proceedings of the ACL 2007 Student Research Workshop*. Prague, Czech Republic, pages 55–60. http://www.aclweb.org/anthology/P/P07/P07-3010.

Yves Scherrer. 2012. Machine translation into multiple dialects: The example of Swiss German. In *7th SIDG Congress - Dialect 2.0*.

Yves Scherrer. 2013. Continuous variation in computational morphology - the example of Swiss German. In *TheoreticAl and Computational MOrphology: New Trends and Synergies (TACMO)*. 19th International Congress of Linguists, Genève, Suisse. http://hal.inria.fr/hal-00851251.

Yves Scherrer and Rambow Owen. 2010. Natural Language Processing for the Swiss German Dialect Area. In *Proceedings of the Conference on Natural Language Processing (KONVENS)*. Saarbrücken, Germany, pages 93–102.

Anne Schiller, Simone Teufel, Christine Stöckert, and Christine Thielen. 1999. Guidelines für das Tagging deutscher Textkorpora mit STTS. http://www.ims.uni-stuttgart.de/forschung/ressourcen/lexika/TagSets/stts-1999.pdf.

Benjamin Snyder, Tahira Naseem, Jacob Eisenstein, and Regina Barzilay. 2008. Unsupervised multilingual learning for POS tagging. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*. Honolulu, Hawaii, pages 1041–1050. http://www.aclweb.org/anthology/D08-1109.

Elisabeth Stark, Simone Ueberwasser, and Anne Göhrig. 2014. Corpus "What's up, Switzerland?". www.whatsup-switzerland.ch.

Milan Straka and Jana Straková. 2017. Tokenizing, pos tagging, lemmatizing and parsing ud 2.0 with udpipe. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*. Vancouver, Canada, pages 88–99. http://www.aclweb.org/anthology/K/K17/K17-3009.pdf.

Oscar Täckström, Ryan McDonald, and Joakim Nivre. 2013. Target language adaptation of discriminative transfer parsers. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Atlanta, Georgia, pages 1061–1071. http://www.aclweb.org/anthology/N13-1126.

Jörg Tiedemann. 2014. Rediscovering annotation projection for cross-lingual parser induction. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*. pages 1854–1864.

Jörg Tiedemann. 2015. Cross-lingual dependency parsing with universal dependencies and predicted PoS labels. In *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)*. pages 340–349.

Jörg Tiedemann, Željko Agić, and Joakim Nivre. 2014. Treebank translation for cross-lingual parser induction. In *Proceedings of the Eighteenth Conference on Computational Natural Language Learning*. Ann Arbor, Michigan, pages 130–140. http://www.aclweb.org/anthology/W14-1614.

David Yarowsky, Grace Ngai, and Richard Wicentowski. 2001. Inducing multilingual text analysis tools via robust projection across aligned corpora. In *Proceedings of the first international conference on Human language technology research*. pages 1–8.

Marcos Zampieri, Shervin Malmasi, Nikola Ljubešić, Preslav Nakov, Ahmed Ali, Jörg Tiedemann, Yves Scherrer, and Noëmi Aepli. 2017. Findings of the VarDial evaluation campaign 2017. In *Proceedings of the Fourth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial)*. Valencia, Spain, pages 1–15. http://www.aclweb.org/anthology/W17-1201.

D. Zeman and Philip Resnik. 2008. Cross-language parser adaptation between related languages. In *Proceedings of the IJCNLP-08 Workshop on NLP for Less Privileged Languages*. pages 35–42.

Daniel Zeman, Martin Popel, Milan Straka, Jan Hajic, Joakim Nivre, Filip Ginter, Juhani Luotolahti, Sampo Pyysalo, Slav Petrov, Martin Potthast, Francis Tyers, Elena Badmaeva, Memduh Gokirmak, Anna Nedoluzhko, Silvie Cinkova, Jan Hajic jr., Jaroslava Hlavacova, Václava Kettnerová, Zdenka Uresova, Jenna Kanerva, Stina Ojala, Anna Missilä, Christopher D. Manning, Sebastian Schuster, Siva Reddy, Dima Taji, Nizar Habash, Herman Leung, Marie-Catherine de Marneffe, Manuela Sanguinetti, Maria Simi, Hiroshi Kanayama, Valeria dePaiva, Kira Droganova, Héctor Martínez Alonso, Çağr Çöltekin, Umut Sulubacak, Hans Uszkoreit, Vivien Macketanz, Aljoscha Burchardt, Kim Harris, Katrin Marheinecke, Georg Rehm, Tolga Kayadelen, Mohammed Attia, Ali Elkahky, Zhuoran Yu, Emily Pitler, Saran Lertpradit, Michael Mandl, Jesse Kirchner, Hector Fernandez Alcalde, Jana Strnadová, Esha Banerjee, Ruli Manurung, Antonio Stella, Atsuko Shimada, Sookyoung Kwak, Gustavo Mendonca, Tatiana Lando, Rattima Nitisaroj, and Josie Li. 2017. Conll 2017 shared task: Multilingual parsing from raw text to universal dependencies. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*. Vancouver, Canada, pages 1–19. http://www.aclweb.org/anthology/K/K17/K17-3001.pdf.